

Combined Virtual Screening Strategies

Alan Talevi^{1,2}, Luciana Gavernet¹ and Luis E. Bruno-Blanch^{*,1}

¹Department of Biological Sciences, Exact Sciences College, Universidad Nacional de La Plata (UNLP), B1900AVV, La Plata, Buenos Aires, Argentina

²Instituto de Investigaciones Físicoquímicas Teóricas y Aplicadas (INIFTA), Department of Chemistry, Exact Sciences College, Universidad Nacional de La Plata (UNLP), B1900AVV, La Plata, Buenos Aires, Argentina

Abstract: The progress in chemical knowledge and synthetic technologies over the last fifty-years has dramatically increased the synthetic accessible chemical entities. Exploration of natural products rich chemodiversity has also expanded the vast chemical universe where medicinal chemist can pursue the identification of new therapeutic agents. Virtual Screening (VS) benefits from computational technology to explore the increasingly vast chemical universe in an efficient manner. The different VS approaches may be characterized by the computational and human time they require, from the highly automated and fast 2D-QSAR ligand-based VS to the more demanding 3D QSAR and target-based (docking) methodologies. Recently, several studies based on the integration of different VS approaches have been proposed, demonstrating that the hit recovery rate may be maintained (or even increased) with a substantial reduction of computing times. Combined virtual screening methodologies usually begin with the least-demanding approaches at the beginning of the VS process and progress to the more accurate, time consuming techniques in the last stages. This review discusses recent 2D/3D QSAR and ligand-based/target-based “synergistic” combinations that allow speeding-up the VS process, permitting accurate and efficient studies on large databases. The impact of the combination of different techniques on the chemical diversity of the compounds retrieved is also discussed.

Keywords: Combined virtual screening, docking, ligand based methodologies, QSAR, pharmacophore.

INTRODUCTION

Faced with the impossibility to exhaustively screen the increasingly vast chemical universe through traditional *in vitro* and *in vivo* assays, researchers in Medicinal Chemistry have developed modern strategies such as Virtual Screening (VS) and High-Throughput Screening (HTS) to explore this huge chemodiversity in an efficient (and bioethically compatible) manner. Although is highly unlikely that *in silico* and *in vitro* tools will ever totally replace *in vivo* assays, ligand- and structure-based VS strategies have been used to filter the ever-growing chemical space, in order to rationally reduce the number of compounds to synthesize/evaluate. With the constant development of new computational methods, VS provides the possibility to analyze a huge number of compounds (including compounds that do not necessarily exist) in a clean, safe and fast manner. Two general and fundamental approaches have been developed: structure-based VS (SBVS) and ligand-based VS (LBVS).

In SBVS, the search of new leads is guided by the knowledge of the 3D structure of the biological target. The procedure, called molecular docking, involves a quantitative analysis of the molecular recognition events between the ligands and the targeted binding site. To this end, the compounds from a virtual library are docked to the structure of the active site of the target protein (by adjusting their conformation to bind to the selected receptor) and the free energies of binding are calculated. Unfortunately, the knowledge

at the molecular level of the drug-receptor interactions that are responsible for the biological recognition is not straightforward, and their simulation on a computer is far from perfect. The influence of the quality of the target information available from experimental data, the unavailability of X-ray or RMN information for some types of molecular targets (e.g. membrane proteins), the selection of the ligand database, the flexibility of the molecules and the receptor, the evaluation and ranking of each ligand active conformation through the scoring functions, the entropy penalty for immobilization of rotatable bonds in binding and the influence of water molecules are some of the subjects in constant discussion [1, 2]. However, it was demonstrated that this kind of VS has made an important contribution to the generation of novel active structures as well to the understanding of protein-ligand interactions [1, 3-8].

Alternatively, ligand-based VS (LBVS) can be performed when there is little or no information available on the molecular target. LBVS methodologies usually involve two fundamental steps: first, structural features significant to elicit biological activity should be studied on a set of ligands which are known to bind the molecular target of interest. Once those critical features have been identified it is possible to apply them in the search of chemical entities that satisfy those requirements in virtual libraries of chemical compounds. In the limit, LBVS may be applied even if a single known-ligand has been identified, through similarity-based VS. LBVS may be classified in two different, essential approaches, which we will denote as high and low dimensional VS throughout this review. High dimensional VS considers the conformational space of the molecules (and sometimes, also, its orientation) and it is, therefore, computationally expensive. In contrast, low dimensional VS is based on a

*Address correspondence to this author at the Department of Biological Sciences, Exact Sciences College, Universidad Nacional de La Plata (UNLP), B1900AVV, La Plata, Buenos Aires, Argentina;
E-mail: lbb@biol.unlp.edu.ar

simpler description of the molecular structure (such as a 2D representation, e.g. a graph, or in some cases, just the molecular formula of the compounds, e.g. when constitutional descriptors or atom counts are used) [9-11].

Plenty debate has arisen in literature discussing whether target-based or ligand-based approaches are more accurate [12-16], and it is usually assumed that SBVS is the alternative of choice provided that quality information on the molecular target is available. A similar discussion can be traced regarding preference on high or low dimensional ligand-based approaches [17-21]. Although intuitively one may believe that those approaches which incorporate more explicit information on molecular structure are more effective, the evidence indicates that sometimes the implicit information captured by simpler descriptions of molecular structure can lead to comparable or even better results (being, at the same time, more cost-effective).

The use of sequential ligand-based filters followed by structure-based filters in VS is justified by the optimization of computational resources to accelerate the processes, increasing the computational requirements of the selected methodologies progressively in each step. When this strategy is applied, the molecular docking usually constitutes one of the last stages, and it is used to provide more information about the interaction between the best ranked ligands and the binding site. Although it is usually assumed that target-based leads to more accurate and reliable predictions (at has been stated in previous paragraphs) recent evidence from systematic studies shows that ligand-based methodologies frequently present best enrichment factors while both SBVS and LBVS performances are highly variable from target to target.

Klabunde and coworkers [22] compared the performance of different virtual screening strategies for identifying biogenic amine-binding G-protein coupled receptors (GPCRs). They selected a library of drug-like molecules and antagonists of four biogenic amine-binding G-protein receptors: alpha1A, 5HT2A, D2 and M1; in order to analyze the ability of the methods to retrieve known antagonist from the database. The screening strategies involved molecular docking (using GOLD [23] and FlexX-PharmTM [24] software), ligand-based pharmacophore (with Catalyst[®] [25] or Feature Trees [26]), 3D-similarity searches (FlexS [27]) and the use of 2D descriptors in QSAR models (CATS [28, 29], NCCS [30], QikProp [31]). The effectiveness of the different methodologies was evaluated by the calculation of the Hit Rates, and the Enrichments factors within the top-scored compounds compared to that of those randomly selected. The Hit Rate is defined as the relation between the number of true hits found in the hit list respect to the total number of compounds in the hit list; and the Enrichment factor (EF) is the HitRate divided by the total number of hits in the full database relative to the total number of compounds in the database. More formally:

$$\text{HitRate} = (\text{Hits}_{\text{sampled}} / \text{N}_{\text{sampled}}) \cdot 100$$

$$\text{EF} = (\text{Hits}_{\text{sampled}} / \text{N}_{\text{sampled}}) / (\text{Hits}_{\text{total}} / \text{N}_{\text{total}})$$

Based in these previous considerations, the authors found that the ligand-based approach offered a better performance than protein-based methodologies, especially when there is sufficient ligand information for the generation of the mod-

els. However, they observed that the docking procedures presented were based in homology models of the protein structures, and the results of the investigation would be affected by the representation of the receptor and by the choice of scoring functions. Additionally, they remarked the significance of the qualitative information provided by the inclusion of structure-based methodologies (mainly related to the molecular recognition) to increase the chances of success in VS.

Another example was provided by Mc-Curdy *et al.* [32], who developed two independent models aimed to identify new tools for the discovery of potential lead molecules for the human kappa opioid receptor (KOP). One of the approaches was based on the automated pharmacophore hypothesis generation (Catalyst software [25]), using a training set composed by salvinorin A analogs. The selected pharmacophore was further validated by randomizing the data, by evaluating the predictive ability on the test set molecules and by incorporating an external set of negative controls. The other model was supported by the virtual generation of the 3D-structure of human KOP receptor. It was derived from X-ray crystal structure of bovine rhodopsin and several checking programs were used for protein structure validation in combination with experimental data. In addition, some derivatives of salvinorin A that are known as KOP receptor agonists were docked with GOLD [23] to correlate their pKi values with the docking scores, with good results. The convergence of the two models was quantified through the correlation of the predicted activities of the test set compounds from pharmacophore model vs their docking scores obtained from human KOP, with acceptable results. The authors stated that the information obtained using both ligand-based and target-based methods will improve the effectiveness and performance in future VS.

Zhang and Muegge have recently published a very interesting systematic comparative study involving seven different drug targets (among them, cyclooxygenase 2, estrogen receptor and HIV-1 protease) [33]. In this study, the authors combined known-ligands of the studied targets with 9,969 drug-like compounds of the MDDR database and performed VS simulations by applying structure based VS with Glide [34], pharmacophore fingerprints (a 3D ligand-based methodology), Daylight fingerprints and atom pair-based similarity measures (2D ligand-based approaches). From the percentages of actives retrieved by each methodology it can be observed that in most cases LBVS outperforms SBVS, while the performance of low-dimensional VS approaches usually compares to (and sometimes, outperforms) that of high-dimensional approaches (Fig. 1).

A similar, notable research was carried out on 11 targets by McGaughey *et al.*, using MDDR and Merck's corporate database as sources of decoys [35]. From the enrichment factors, it was observed that topological similarity (the simplest approach) outperformed 3D ligand-based and docking methodologies in most cases. However, the authors proved that more complex methodologies tend to retrieve more diverse compounds; in this sense, 3D ligand-based methodologies may be considered generally better, since they present higher enrichment factors while the diversity of the retrieved compounds is similar to that of structure-based techniques. Nevertheless, it may be signaled once again that

Fig. (1). Comparison of different VS methodologies on seven molecular targets, from the work from Zhang and Muegge. Note that 2D and 3D-ligand based VS methodologies usually outperform target-based VS in terms of enrichment factors. G represents Glide, P1 represents pharmacophore fingerprints from a single 3D conformations generated by using Corina, P50 represents a fingerprint generated by combination the pharmacophore fingerprint information from 50 conformations generated by OMEGA, D represents Daylight fingerprints and AP represents atom pairs.

docking methodologies present an additional virtue: the qualitative information given by the possibility to predict the binding mode of the ligand. In synthesis, 2D ligand-based approaches appear to be generally better if one takes the number of actives retrieved and the computational cost as the only performance criteria, while more complex approaches are better if the diversity of the retrieved structures or the information provided on the molecular recognition events are considered. Thus, each methodology presents its own pros and cons, and there is no clear superiority of one over the others.

Consistently with these results in the last few years many successful VS applications combined either different ligand-based approaches or structure-based with ligand-based techniques in complex VS protocols. Simultaneously, plenty effort is being invested in the search of consensus scoring to merge results from highly disparate VS methods [36]. These works support the idea that different VS alternatives should not be regarded as mutual exclusive but integrated in a synthetic, synergistic way instead. However, is there any information loss related to these integrated strategies?

In this review we will focus on these recent applications of combined VS approaches. We have divided the review in three sections, covering *combined ligand-based and structure-based approaches*, *combined high and low dimensional ligand based approaches* and *combinations of different structure-based VS techniques* (cascade docking). To help the understanding of the more complex VS protocols we have included schemes of the combined VS processes. Finally, we will draw some general conclusions underlining the advantages, disadvantages and perspectives of the different combined VS protocols.

COMBINED LIGAND-BASED AND STRUCTURE-BASED VS APPROACHES

Recently, Hu *et al.* combined a learning support vector machine model with multiple conformational docking for the

discovery of *Yersinia* protein kinase A (YpkA) inhibitors [37]. A set of 364 known kinase inhibitors assembled from literature or from ligand-protein complexes in the Protein Data Bank, covering diverse chemical scaffolds, was seeded into 4,220 compounds randomly selected from MDDR database. The model derived from 276 3D descriptors computed with ADMET/Predictor [38] correctly classified 319 out of 364 active compounds in the training set. The model was externally validated, correctly classifying about 70% of the active compounds from the test set (false positives rate, however, is not reported for neither the training nor the test set). The database was filtered with this model, screening an in-house compound collection of more than 2 million compounds and obtaining a kinase focused library of 200,000 structures. Since the structure of YpkA is unavailable, homology models were constructed from mitogen-activated protein kinases (MAPK) considering the residues near the ATP binding site. Multiple conformational docking was then applied to search for YpkA inhibitors using FlexE and a total of eight conformers of YpkA [39]. The 200,000 compounds of the kinase-focused library were docked and ranked according to FlexX score. Later, consensus scoring was applied for the FlexX docking complexes; the top 5% compounds were re-ranked with the X-Score [40]. The top 1,000 compounds were visually inspected in terms of overall fit, interactions with the binding site, structural complexity and diversity, selecting 45 compounds for biological evaluation. Seven of them (15%) showed complete inhibition at 225-450 μM ; further evaluation proved three of them have IC_{50} values below 10 μM , while the remaining presented IC_{50} values below 50 μM . Interestingly, the selected compounds present diverse scaffolds and three of them showed selectivity towards YpkA (5 to 10 fold better inhibition on YpkA than in MAPK).

Steindl *et al.* applied a combined VS protocol which included pharmacophore fitting, docking and ligand clustering through Principal Components Analysis (PCA), to select new inhibitors of the coat protein of the human rhinovirus (HRV) (Fig. 2) [41]. In the first stage of the VS process

structure-based pharmacophore model was proposed from the crystal structure of the HRV coat protein complexed with an inhibitor, using Catalyst [25]. The initial pharmacophore hypothesis was validated in a simulated screening of 22 HRV coat protein inhibitors seeded among approximately 50,000 compounds from Derwent WDI, and applied later in the screening of Maybridge database, considering multiple conformers for the approximately 60,000 structures in the database. Since the first pharmacophore guess proved to be far too little restrictive (selecting 14.1% of Derwent WDI DB and 15% of Maybridge DB), it was refined through two subsequent, new hypothesis; the last of them retrieved only 10 compounds from Maybridge DB. The docking tool LigandFit included in the Cerius software [42] was then applied as a second filtering criterion. Several conformations of the docked molecules were taken into account using a Monte Carlo algorithm for conformation generation, while the protein was kept rigid. All the seven scoring functions included in Cerius were originally considered, as well as a consensus score approximation. Simulated docking VS was applied to select suitable docking parameters and scoring conditions. To this end, 10 known HRV coat protein inhibitors were seeded among 990 presumably inactive molecules generated with the software iLib diverse and it was checked which docking methodology ranked the active compounds in high positions retrieving the hits in correct orientations and conformations compared to the information of the X-ray structures [43]. Finally, PCA-based clustering was applied to verify that the compounds selected from Maybridge database were clustered near 10 known active ligands using 34 Cerius descriptors, among the several descriptors related to ADME properties. The combined results of these VS approaches allowed the selection of six candidates, which were submitted to biological testing. All six compounds inhibited viral growth and showed activities in the micromolar range, with one of them having an IC_{50} value as low as 4.3 μ M. From the observation of cytotoxicity and difficulties related to solubility in some of the compounds, the authors have pointed the need of more cautious estimation of these properties in future studies.

A similar VS protocol based on the combination of pharmacophore queries and docking was performed by Rastelli and collaborators [44], in the search of inhibitors of *Plasmodium falciparum* bifunctional dihydrofolate reductase-thymidylate synthase (DHFR-TS), pointed to the identification of new inhibitors with activity against antifolate resistant malaria. Six Catalyst pharmacophores were built upon the protonated and neutral structures of cycloguanil, pyrimethamine and WR99210. Since some studies indicate that resistance is due to accumulation of point mutations in the DHFR domain of the enzyme, as stated by the "steric constrain hypothesis" [45], one of the two methyl substituents at the C2 position of cycloguanil was excluded from the overall pharmacophore volume, to avoid steric clash in the cycloguanil resistant DHFR. The pharmacophores were applied in the search of new inhibitors through the ACD database (230,000 compounds) yielding 4,061 candidates for docking. A pKa estimation was performed to determine if the compounds selected by the protonated pharmacophores could be protonated at physiological pH. It is worth mentioning that the diversity of this focused library was assessed by way of the Tanimoto coefficient, demonstrating a better

molecular diversity in the focused library than in the ACD database. This is in good agreement with the statements of the introduction of this review: 3D ligand-based and docking VS methodologies tend to retrieve diverse scaffolds from the database, especially when the 3D ligand-based approach is based in generic functions such as hydrogen bond donors and acceptors, hydrophobic groups and molecular volumes. Docking was performed with DOCK3.5 [46], and the best 500 ranked compounds were clustered into chemical families, excluding those molecules related to classic antifolates. The best ranked compounds within each family were tested against the recombinant wild type and antifolate resistant *Plasmodium falciparum* DHFR. 12 new DHFR inhibitors were identified, with inhibitory activity in the micromolar range and chemical structure unrelated to those of classical antimalaric drugs. Although they are all quite less effective than cycloguanil and pyrimethamine, it is worth emphasizing that, unlike these two drugs, their activity on mutant DHFR is similar or sometimes higher than in the wild type DHFR.

Guo and co-workers employed VS to identify novel hits for 3-Hydroxy-3-methylglutaryl-coenzyme A reductase (HMGR) inhibitors [47]. The investigation started with CoMFA studies on a training set of statins and statin derivatives. Two molecular alignments were used: (1) by fitting the corresponding atoms of the compounds with a template (rosuvastatin inhibitor, from which crystal structure of enzyme-complex was determined) and (2) by using the molecular docking active conformation (Flex-Pharm software [24]), taking into account protein structural variability. The two models generated similar results, and the contour plots obtained matched well with the active site topological surface generated from crystal structure of HMGR active site. According to the results obtained, a new pharmacophore was generated by the fingerprints hypothesis technique (Tuplet software [48]), which was used in VS of 41,393 entries pre-selected according to Lipinski rules. A set of 4,138 structures that satisfied the pharmacophore query were docked into the HMGR active site and were used for another CoMFA modeling. Finally, the 2,291 best ranked hits obtained from docking and/or CoMFA were re-docked taking into account protein structural variability using FlexX. The investigation led to eight potential HMGR not statin-like structures with high pIC_{50} predicted values.

Another example was presented by Spalluto *et al.* [49]. The authors applied a combined approach for the improvement of the selectivity at human A_3 adenosine receptor antagonists. A set of 106 pyrazolo-triazolo-pyrimidine derivatives previously synthesized and evaluated were subjected to high throughput molecular docking with MOE [50], using an improved representation of the human A_3 receptor obtained by homology modeling from the X-ray crystal structure of bovine rhodopsin. They did not find a good correlation between experimental binding affinities and theoretical binding energies based on scoring functions, but the docking procedure allowed the identification of structural features to define a novel pharmacophoric pattern. Consecutively, the best docking scored conformations were used as input for CoMFA studies and the contour plots obtained were coherent with the results originated by docking procedures. On the basis of both the target-based pharmacophore and the CoMFA model 17 new derivatives were synthesized and evaluated. The predicted pKi values were very close to the

Fig. (3). Flowchart of the methodologies proposed by Spalluto *et al.*

experimental ones. In particular, one of the derivatives (4-[3-(2-furan-2-yl-8-methyl-8H-pyrazolo[4,3-e][1,2,4]triazolo[1,5-c] pyrimidin-5-yl)ureido]phenyl)acetic acid ethyl ester) presents activity in a sub-nanomolar range (0.82nM); which was properly predicted by the regression model. Fig. (3) represents the combined design strategy used by the authors.

Fig. (4) presents a scheme of the combined VS protocol used in the very interesting work from Luzhkov *et al.* [51]. The authors applied a 2D/3D indirect VS approach in combination with a direct, structure-based method, in the search of inhibitors of the Flavivirus mRNNA cap methyltransferases (MTases). Remarkably, they conducted their study in an iterative way, using the best structure of the first VS round as the query entrance for a second VS round. S-adenosyl-L-homocysteine (AdoHcy), a co-product of the methyl transfer, was first applied as search query in a 2D similarity search of a database of 2.1 million commercially available compounds. Since the use of unmodified AdoHcy only retrieved chemically protected adenosinetriphosphate analogues, the adenine group in AdoHcy was replaced by a phenyl moiety. Compounds with molecular weight above 250 g/mol and Tanimoto index above 0.40 were retained in the 2D similarity search, leading to 1,671 hits. Independently, a ligand-based VS approach using a pharmacophore model was conducted, resulting in 895 hits. The geometries of the compounds extracted from the pharmacophore and similarity searches were optimized, and the structures were then prepared for a docking stage, considering possible tautomers, stereoisomers, ring conformations and ionization states, resulting in 7,836 total structures. Docking was performed with GOLD [23], taking into account both local and global docking (docking regions of 13 and 22 Å³ spheres around the binding site of the methylating cofactor S-adenosylmethionina –AdoMet- in that order). The authors included the use of global docking to allow the possibility of identifying multiple binding sites, as well as indicating the

selectivity of a given ligand for the primary target site. In order to select, latter, a diversity of scaffolds for biological testing, the 7,836 structures from the LBVS were split in 20 structurally diverse ligand clusters using hierarchical clustering based on atom pairs and fingerprints. These structures were submitted to local docking using the program settings for faster optimization. Depending on the size of each cluster, the top-ranking 5-20 hits from each cluster were merged into a set of 230 ligands, which was re-docked using programs settings for accurate search, considering both local and global docking. Note that other studies which combine docking techniques characterized by different computational cost and accuracy are described later in this review, in the correspondent section. Nine compounds were selected for biological testing; most of the compounds resulted weak inhibitors of mRNA cap methyltransferase. The compound which showed the best activity (IC₅₀ = 60 μM) was used as query for a 2D substructure search; this second VS round produced 3,225 structures, which were once again clustered and screened through local and global docking, leading to 6 new compounds, 4 of which presented weak inhibitory activity at biological testing.

Another interesting recursive approach was performed by Hammond *et al.* in order to speed up the SBVS process [52]. They considered the QSAR solutions based on a combination of global inductive and conventional QSAR descriptors, to create quantitative models that enable evaluations of hypothetical docking scores for defined protein-ligand systems. The target protein was initially the human sex hormone binding globulin (SHBG), since the discovery of nonsteroidal SHBG inhibitors constitutes an attractive target for hormone-replacement therapies to combat diseases related to steroid insufficiency. Initially, the authors applied some conventional pre-docking filters to the NCI database (such as Lipinski's rules for drug-likeness criteria) and the QSAR inductive parameters were calculated for the resulting set

Fig. (4). Scheme of the VS protocol proposed by Luzhkov *et al.*

(composed by 89,941 nonsteroidal structures). The resulting set of 90,184 entries composed by the NCI drug-like compounds and structures from an extended steroids benchmark set previously collected [53] was clustered and a significant set of compounds (that covers the chemical space of the database) was docked into the active site of the crystal structure of the protein corresponding to 1KDM in Protein Data Bank using Glide software [34]. The best docking scores obtained were used as the dependent variable to construct a linear QSAR model, which gave out hypothetical scores for the remaining undocked structures. After that, the compounds with worst predicted scores were removed of the undocked set of structures and the docking process was repeated until all entries of the docking database were processed. Fig. (5) illustrates the general scheme of the procedure. This methodology allows the removal of obvious non-binders before the docking process, reducing the computational time and speeding up the process. An exhaustive analysis of the capacity of recovering effective binders with different settings (related to the sample size and rejection percentage) indicated that in all cases the progressive docking permitted an exponential recovery of the true positive predictions, and the number of false negatives predictions was not very high in most cases. The optimal settings found were 10,000 molecules for the initial sampling step and a 20% rejection criteria, which allowed filtering out 38,301 of 90,184 molecules. The procedure was then validated on other targets, showing that it recovered 80-100% of potential docking hits and saving up 40% of the Glide processing time. Finally the method was applied to another crystal structure of SHGB (PDB code: 1D2S), selected to complement the previous studies based on 1KDM. Experimental testing considering the top docking resultant compounds for 1KDM and 1D2S served to identify four compounds with binding affinities in the low nanomolar range.

It is worth mentioning that this approach was based on the assumption that the binding affinities are not considerably related to the molecular orientation in the active site, but the applicability of other systems remains to be investigated. This subjacent general idea of minimizing the computational costs and time involved in the structure-based screenings by means of a QSAR model derived from the docking scores of a small set of known ligands was previously considered by other authors. In preceding investigations it was shown that the use of the Bayesian classifiers trained on docking scores of experimentally known binders can be used for virtual screening [54-57]. In view of that, Filikov and coworkers also developed a surrogate docking method to pre-rank compounds to regular docking [58]. In this case the target protein was the cyclin-dependent kinase 2 (CDK2) and two sources of compounds were used: the NCI library and a combinatorial library composed by a set of 2,6,9-trisubstituted purines, a chemical family known to inhibit CDK2 as well as other kinases. They optimized the descriptors used in the QSAR model, the size of the training set and the cutoff score for binders using Bayesian classifier. Then, with these optimized values determined, the surrogate docking method using Bayesian classification and a Radial Basis Function (RBF) was performed. It was found that both tools provide significant enrichment docking hits, but the Bayesian method is much faster and identifies structural features found in binders and in non binders. Finally, the Bayesian method was applied to

a set of 205 compounds with known affinities to estradiol receptor, and it was confirmed that in this case the ability to enrich a database with experimental binders using surrogated docking is comparable to the one obtained by docking itself.

According to the examples mentioned before, in sequential VS the docking methodologies comes out as one of the last stages of the processes mainly due to its computational cost. However, Baker and co-workers performed a different approach for the design of new autotoxin (ATX) inhibitors, in view of the limitations found for the available ligands [59]. The current known ATX inhibitors exhibit a low structural diversity, and they present a high grade of flexibility that forbid the definition of a single pharmacophore model. Additionally, they fail to meet the Lipinski's rules that characterize drug-like compounds. Consequently, the authors applied SBVS methodologies in order to identify new inhibitors with drug-like physicochemical properties (and to gain diversity for the creation of a suitable set of compounds for QSAR model). The ATX catalytic domain was developed by homology modeling using the crystal structure of a bacterial nucleotide pyrophosphatase/phosphodiesterase (NPP) enzyme. MOE was used for docking studies [50], in presence and absence of a restraint related to an anionic atom of the ligand near one catalytic residue (T210). Better results were obtained for the last case. Docking candidates were selected using the online search tool www.hit2lead.com (Chem-Bridge), considering compounds with anionic groups or metal ligating functions, as well another structural features (such as the presence of an additional aromatic ring, and logP/molecular weight restrictions). 500 compounds were docked, and 94 of them were experimentally evaluated. The results showed that 20% of the tested structures were active (with 20% of enzyme inhibition or more, at a concentration of 10 μ M). In addition to the contribution to the knowledge of the enzyme-ligand interactions, docking studies provided the basis for binary QSAR models. They were developed using 97 structures identified previously using SBVS or taken from literature. The first model considered the compounds obtained from the database without further processing. The second model includes explicit hydrogen atoms, ionization states expected under assay conditions and geometry optimization. Most of the descriptors were calculated with MOE software [50]. Validation results showed that Model 1 presented several advantages over Model 2, such as the identification of a greater number of active compounds and the recognition of the most effective structures, with no need of processing the database.

COMBINED 2D/3D LIGAND-BASED VS APPROACHES

Even though most of the recent articles that present combined VS approaches focuses on integration of LBVS and SBVS methodologies, we have found some examples of "pure" combined ligand-based protocols. However, it is worth emphasizing that some of the ligand-based stages from the combined ligand/target based approaches reviewed in the previous section do combine 2D and 3D ligand-based VS approaches [41, 51]. Considering that the pros and cons of each particular methodology have been mentioned and that we have stated that no method is clearly superior over the others, the exclusion of any of these methodologies does not seem as a natural option, except in those cases where some

Fig. (5). General scheme for the methodology developed by Hammod *et al.* The recursive cycle is illustrated by the dashed lines and the non-colored objects, and the preparation cycle is illustrated by the solid lines and grayed object.

specific limitation prevents from using a particular approach (for example, lack of the 3D structure of the target or impossibility of obtaining a homology model), although it is worth questioning if 2D ligand-based approaches should be used serially or in parallel when integrated to 3D LBVS and SBVS, provided that low dimensional ligand-based techniques tend to select less diverse hits and therefore might jeopardize the ability of 3D ligand-based or structure-based methods to select novel scaffolds. This will be further discussed in the conclusions section.

In the absence of the 3D structure of sodium channels linked to epilepsy, Gavernet *et al.* performed a combined 2D/3D VS approach that includes the subsequent use of a discriminant function based on topological descriptors from the program Dragon [60], general ADME filters such as Lipinski's rule of five and Veber's rules for estimating drug bioavailability and the optimal log P value for a compound to diffuse passively through the blood-brain-barrier, and, finally, fitting to a Sybyl pharmacophore [61]. From more than 10,000 structures from InterBioScreen database, the 2D/ADME filters yielded a total of 56 structures, which were fitted to the pharmacophore. 4 structures with an energy difference of less than 4 kcal/mol between the minimum energy conformation and the active conformation defined by the pharmacophore were selected as anticonvulsant candi-

dates. One of them was tested in the MES test and proved active in mice at 30 and 100 mg/kg, the lower doses of Phase I of the NIH Anticonvulsant Drug Development Program.

In their quest for a novel NPY5 receptor antagonist to develop anti-obesity medications, a research group from La Roche performed a VS campaign incorporating topological information in a Catalyst 3D pharmacophore, so that the ability of 2D similarity methodologies to retrieve hits with strong binding affinity were merged with the potential of 3D pharmacophores to yield structurally diverse scaffolds [62]. The VS of Roche compound repository retrieved 632 molecules from which 31 presented IC_{50} below 10 μ M. One of the compounds presented very low Tanimoto similarity (0.23) when compared to the query structure and was free of patent claims.

CASCADE DOCKING

When applying target-based VS methodologies, one strategy is to take a direct approach to the screening, docking the compounds of the database to the target using a certain, single docking tool. An alternative, increasingly popular strategy is the use of several fast computational techniques first, yielding an enriched, focused subset of compounds that are then docked through more accurate, computationally demanding tools.

Mukherjee *et al.* have implemented a 3-stage cascade docking aimed to the selection of the Chymotrypsin-like cysteine protease of SARS coronavirus, looking for just shape complementarity at the early stages and focusing on binding site interactions and geometrical quality of the binding pose later [63]. Departing from Asinex Platinum collection (120,000 compounds) they performed a pre-filtration with ADME filters retaining only 32,000 compounds. Using GOLD [23], half of those structures were disregarded through GOLD 7-8 times (the generation of a single docking pose is reduced 7-8 times compared to the standard mode), and half of the remaining were discarded with GOLD 2 times. The final screening with GOLD standard led to 500 hits, which were submitted to clustering analysis, retaining 100 compounds comprising the top-ranked molecule from every cluster. 27 compounds were selected for biological testing through visual inspection. One of them presented a IC_{50} of 18.2 μ M. Interestingly, a second VS protocol incorporating pharmacophore matching stages to the previous protocol yielded a different set of 81 candidates, one of which presented a IC_{50} of 17.2 μ M, which highlights the already stated fact that different VS approaches yield different results and that they should be regarded as complementary in nature. Note that similar multi-stage cascade docking protocols had been successfully applied previously by the same group, aimed to the identification of parasitic cysteine protease inhibitors [64, 65]. However, the pharmacophore approach had not been incorporated in the VS protocol of those previous studies. Other studies showed that inclusion of a pharmacophore-based filter can significantly improve the probability of identifying potential hits [66, 67]. A very interesting contribution from Maiorov and Sheridan shows that the opposite sense may also be true: the enrichment ratio obtained with the use of an approach combining pharmacophore and a simplistic docking algorithm can be substantially improved with more demanding docking tools [68]. Through a systematic study on 20 proteins the authors proved that the enrichment ratio ICM-Dock to FLOG is between 2 and more than 8 in many cases [69, 70].

In a revealing study, Miteva *et al.* performed combined fast rigid-docking approach (FRED) with more complex and computationally demanding flexible approaches in DOCK and Surflex [71-73]. Comparison of the results of the multi-stage methods to those of the more complex approaches applied alone, on four molecular targets, showed that, sur-

prisingly, the performance of the combined rigid and flexible approaches was better than that of the flexible approaches alone, not only in terms of computational time but in terms of enrichment factors as well. The authors suggested that the more flexible approaches used by Dock or Surflex succeeded to fit some molecules into the binding pockets with favorable contact scores while they had relatively low shape complementarity. In other words, difficulties regarding ligand flexibility some times may lead to worse results than in those cases in which simpler filters are included at earlier stages. A good example of this approach may be found in the work from Wang *et al.* on HIV-1 reverse transcriptase, where after an initial pharmacophore filtering of 150,000 compounds, the authors applied multi-conformational rigid docking, solvation docking and, finally, MD simulations on the 30 best hits from the former stages to sample both the ligands and target conformational space [74].

CONCLUSIONS

Although throughout the years there has been plenty debate and studies aimed to define which of the many VS approaches is the best, the question remains with no conclusive answer but the fact that there might not be a definitively superior methodology. Each approach has its own advantages and drawbacks (a brief summary of each approach can be found in Table 1, and the choice of one or other depends on the particular scenario faced by the Medicinal Chemist and the aim of the study. If one is looking for an efficient methodology, ligand-based applications might be the best choice. Among them, 2D methodologies tend to present better enrichment factors. Target-based, direct approaches should be the alternative of choice whenever information regarding the interaction between the drug and the molecular target is sought. Additionally, the least efficient approaches such as 3D ligand-based and docking seem to have advantage over the 2D approaches when new, highly diverse scaffolds are being sought.

In the last five years many studies have chosen not to choose a particular, single VS approach but a synergistic, rational, synthetic combination of different approaches instead, clearly aiming to take advantage from the best of each particular methodology. A summary of the articles commented in this review (from systematic comparisons of ligand-based and structure-based methodologies to cascade docking) can be found in Table 2. With the exception of the

Table 1. A Summary of the Advantages and Drawbacks of Individual VS Approaches

VS Approach	Advantages	Disadvantages
2D ligand-based (2D similarity, 2D QSAR)	<ul style="list-style-type: none"> • Fast • High enrichment factor • High reproducibility: especially in the case of similarity-based 2D approaches, there are few decisions to be made by the operator. 	<ul style="list-style-type: none"> • 2D LBVS tends to select less diverse hits than more complex approaches
3D ligand-based (3D QSAR, pharmacophore based techniques)	<ul style="list-style-type: none"> • High dimensional VS approaches tend to select more diverse, novel scaffolds than 2D methodologies 	<ul style="list-style-type: none"> • Computational cost: the VS process usually involves conformational analysis.
Structure-based	<ul style="list-style-type: none"> • SBVS tends to select novel and highly diverse scaffolds. • Docking gives information on molecular recognition events. 	<ul style="list-style-type: none"> • Computational cost. • Results critically depend on operator choices (pre-processing of the target, scoring function, etc.).

Table 2. A Summary of the More Prominent Articles Commented in Each Section of this Review

Comparison Between Methods			
Methods	Reference	Target	Comments
Molecular docking Pharmacophoric patterns 3D-similarity searches 2D-QSAR models	Klabunde <i>et al.</i> [22]	G-protein receptors	A comparison between several methodologies indicated that the better performances were achieved by ligand-based models in this system.
Pharmacophoric patterns Molecular docking	Mc-Curdy <i>et al.</i> [32]	Kappa opioid receptor	An acceptable convergence was found between the pharmacophore predicted activities and the docking scores.
Molecular Docking Pharmacophoric patterns Pair-based similarity approaches	Zhang and Muegge [33]	CDK2, COX2, estrogen receptor, neuraminidase, HIV-1 protease, p38 MAP kinase, thrombin	In most cases ligand-based approaches outperform structure-based Virtual screening in studied the systems.
Molecular Docking 3D ligand-based approach Topological similarity	McGaughey <i>et al.</i> [35]	Carbonic Anhydrase Protein Kinase Cyclooxygenase Dihydrofolate Reductase Estrogen receptor HIV protease Reverse Transcriptase Neuraminidase Thrombin Thymidylate Synthetase Tyrosine phosphatase	Topological similarity gives the best performance, however the other methodologies retrieve more diverse compounds.
Combined Ligand-Based and Structure-Based VS Approaches			
Methods	Reference	Target	Comments
3DQSAR methods Molecular Docking	Hu <i>et al.</i> [37]	<i>Yersinia</i> protein kinase A	Identification of inhibitors with diverse structural scaffolds that present selectivity against the enzyme.
Pharmacophoric patterns Molecular Docking Ligand clustering	Steindl <i>et al.</i> [41]	human rhinovirus	Example of the use of sequential methodologies to the detection of six active compounds at the micromolar range.
Pharmacophoric patterns Molecular Docking	Rastelli <i>et al.</i> [44]	<i>Plasmodium falciparum</i> bifunctional dihydrofolate reductase-thymidylate synthase	Twelve new inhibitors (micromolar range) were identified using sequential methodologies. The compounds were structurally different to those of classical inhibitors.
CoMFA Pharmacophoric pattern Lipinski rules Molecular Docking	Guo <i>et al.</i> [47]	3-Hydroxy-3-methylglutaryl-coenzyme A reductase	Identification of eight potential inhibitors with not statin-like structures and high pIC ₅₀ predicted values.
Molecular Docking Pharmacophoric pattern CoMFA	Spalluto <i>et al.</i> [49]	human A ₃ adenosine receptor	Discovery of one inhibitor in a sub-nanomolar range correctly predicted by the models.
2D similarity search Pharmacophore model Molecular Docking	Lushkov <i>et al.</i> [51]	Flavivirus mRNA cap methyltransferases	Iterative approach that evolve to the discovery of 4 compounds with weak inhibitory activity
QSAR Lipinski rules Molecular Docking	Hammond <i>et al.</i> [52]	human sex hormone binding globulin	Example of another recursive model to speed up the docking methodologies using QSAR. Four compounds with binding affinities in the low nanomolar range were found.
QSAR Molecular Docking	Filikov <i>et al.</i> [58]	cyclin-dependent kinase 2	The analysis of the ability to enrich a database with experimental binders showed that the performance of QSAR methods is comparable to the one of docking itself.
Molecular Docking QSAR	Baker <i>et al.</i> [59]	autotoxin	In this example a QSAR model was developed using active structures found previously in docking studies. 20% of the compounds selected using Docking inhibit the enzyme, while the QSAR model was 37% accurate in the selection of active compounds for screening.

(Table 2) contd.....

Combined 2D/3D Ligand-Based VS Approaches			
Methods	Reference	Target	Comments
2D-discriminant function 2D/ADME filters Pharmacophore model	Gavernet <i>et al.</i>	sodium channels	Identification of novel anticonvulsant compounds.
2D-similarity methodologies Pharmacophore model	Mukherjee <i>et al.</i> [62]	NPY5 receptor	The investigation 31 diverse structures with IC ₅₀ below 10 μM.
Cascade Docking			
Methods	Reference	Target	Comments
ADME filters Molecular Docking Pharmacophore model	Mukherjee <i>et al.</i> [63]	Chymotrypsin-like cysteine protease	Identification of active compounds with Docking. The protocol that incorporates the pharmacophore matching stages yields a more active compound (IC ₅₀ =17.2)
Pharmacophore model Molecular Docking	Sheridan <i>et al.</i> [68-70]	20 proteins	More demanding docking tools such ICM-Dock improves the FLOG enrichment ratio.
Molecular Docking	Miteva <i>et al.</i> [71-73].	Several targets	The performance of the combined rigid and flexible docking stages was better than that of the flexible approaches alone in the studied systems.
Pharmacophoric patterns Molecular Docking Molecular Dynamic simulations	Wang <i>et al.</i> [74]	HIV-1 reverse transcriptase	Example of another docking cascade that involved multi-conformational rigid docking, salvation docking and ended with MD simulations.

work from Luzhkov *et al.*, it can be noticed that combined VS protocols tend to include less costly approaches at the first stages of the VS protocol, while reserving the most demanding methods for the last stages, when the original large compound databases have been reduced to a focused library of manageable size. Compressively, much emphasis is being given to the fact that the candidates selected for enzymatic/biological testing by the end of the VS protocol should be structurally diverse, representatives of as many chemical families as possible, in order to identify new active, patent-free, scaffolds. To this end, complex VS protocols usually include, between the different VS stages, a clustering step to select the best-ranked structures among each chemical family and similarity calculations to assess the diversity of the hits of a particular VS stage. Bearing in mind that 2D methodologies tend to select similar hits, the implementation of strategies to assure structural diversity is especially important whenever 2D VS is combined with either 3D ligand-based approaches (such as pharmacophore fitting) or docking. Furthermore, it is worth questioning if the combination of 2D similarity based approaches with 3D approximations should be performed in the form of serial, subsequent filters (with the most efficient, 2D filters at the first stages of the VS) or in parallel, merging candidates from low and high dimensional ligand-based approaches in a single pool. If interest in high dimensional approaches is founded in the expectancy of a set of highly diverse hits, then the use of a 2D similarity based filter in a previous stage might be counterproductive if one bears in mind that the latter tend to select a narrower range of scaffolds. In that case, a good choice might be the parallel, independent use of 2D and 3D ligand-based approaches previous to docking (as in the work of Luzhkov *et al.*) or the use of efficient ligand-based 3D meth-

ods or computationally inexpensive docking approaches previous to more accurate, computationally demanding target-based procedures (which, as has been showed by several of the reviewed studies, not only improves computational times but might identify new candidates as well). In contrast, if, for example, the search is aimed to retrieve a number of structurally related hits to establish structure-activity relationships or one is interested in gaining information on the chemical interactions between hits from a 2D VS and the molecular target, then the use of 2D filters in the first steps of the VS protocol is legitimate. Fig. (6) summarizes different integrated VS protocols designs, together with the advantages and disadvantages of each scheme.

It is also worth investigating the possible advantages of alternative designs of integrated ligand-based/structured-based protocols, beside the classic “less costly filters at the beginning, more costly filters at the end” e.g. if a novel scaffold has been identified through high dimensional ligand-based or target-based approaches, it might be attractive to include simpler, similarity based VS stages at the end of the protocol, in order to identify new compounds related to the novel scaffold but, for example, with different, more favorable ADMET properties.

Finally, it is interesting to underline once more that some studies rise evidence on the fact that sometimes the use of simpler approaches in earlier stages of a multi-stage VS approach is not only functional in terms of timing, but also in terms of improved enrichment factors, probably due to the fact that more complex tools, require more expertise from the human operator to define the program’s parameters settings most adequate to the problem at hand, while conformational sampling sometimes adds significant noise to the re-

Fig. (6). Three basic different schemes of integrated LBVS and SBVS approaches, and a summary of the correspondent advantages and disadvantages. A dotted line indicates that one step can or can not be in the protocol (for example, there might be an integrated VS protocol which only incorporates serial low and high dimensional LBVS approaches, or low dimensional LBVS and SBVS). Iterative cycles (using the results of a cycle as initial query for the next) are also possible. Some studies integrate more than one of these simple protocol choices (see the work from Luzhkov *et al.* as example: parallel high dimensional and low dimensional LBVS is followed by SBVS which is followed by low dimensional similarity-based LBVS).

sults. As demonstrated in the case of combinations of rigid and flexible docking tools, the application of simpler filters may compensate the sometimes excessive flexibility of the flexible approaches. As exemplified by many of the cited work, the strategy used in each step from a multi-step VS protocol should always be validated, if possible, with a simulated screening, seeding a number of known ligands among a large library of presumably inactive compounds, to exclude

the use of too flexible VS tools with poor ability to discriminate between the active and inactive compounds. When simulated screening of this kind is performed, one should bear in mind that the enrichment factor dramatically depends on the kind of decoys used (universal, drug-like, mimetic or modeled decoys) as brilliantly stated recently by Nicholls [75].

ACKNOWLEDGEMENTS

Figs. (1, 3, 5) are reproduced under permission of the American Chemical Society. We are thankful to Drs. Ingo Muegge and Artem Cherkasov for gently providing us high resolution images for Figs. (1, 5).

A. Talevi is a fellowship holder of Consejo Nacional de Investigaciones Científicas y Técnicas de la Republica Argentina (CONICET). L. Bruno-Blanch and L. Gavernet are members of the Facultad de Ciencias Exactas, Universidad Nacional de La Plata. L. Gavernet thanks CONICET. This research was supported in part through grants from Agencia Nacional de Promoción Científica y Tecnológica (PICT 06-11985/2004), CONICET, and Universidad Nacional de La Plata, Argentina.

REFERENCES

- [1] Klebe, G. Virtual ligand screening: strategies, perspectives and limitations. *Drug Discov. Today*, **2006**, *11*, 580-594.
- [2] Kitchen, D.B.; Decornez H.; Furr J.R.; Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature Rev.*, **2004**, *3*, 935-949.
- [3] Lyne, P.D.; Kenny, P.W.; Cosgrove, D.A.; Deng, C.; Zabludoff, S.; Ashwell, S.; Wendoloski, J.J. Identification of compounds with nanomolar binding affinity for checkpoint kinase-1 using knowledge-based virtual screening. *J. Med. Chem.*, **2004**, *47*, 1962-1968.
- [4] Barril, X.; Brough, P.; Drysdale, M.; Hubbard, R.E.; Massey, A.; Surgenor, A.; Wright L. Structure-based discovery of a new class of Hsp90 inhibitors. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 5187-5191.
- [5] Foloppe, N.; Fisher, L.M.; Howes, R.; Potter, A.; Robertson, A.G.S.; Surgenor A.E. Identification of chemically diverse Chk1 inhibitors by receptor-based virtual screening. *Bioorg. Med. Chem.*, **2005**, *14*, 4792-4802.
- [6] Rummey, C.; Nordhoff, S.; Thiemann, M.; Metz G. *In silico* fragment-based discovery of DPP-IV S1 pocket binders. *Bioorg. Med. Chem. Lett.*, **2006**, *16*, 1405-1409.
- [7] Cavasotto, C.N.; Ortiz M.A.; Abagyan, R.A.; Piedrafita F.J. *In silico* identification of novel EGFR inhibitors with antiproliferative activity against cancer cells. *Bioorg. Med. Chem. Lett.*, **2006**, *16*, 1969-1974.
- [8] Kellenberger, E.; Springael, J.-Y.; Parmentier, M.; Hachet-Haas, M.; Galzi, J.-L.; Rognan, D. Identification of nonpeptide CCR5 receptor agonists by structure-based virtual screening. *J. Med. Chem.*, **2007**, *50*, 1294-1303.
- [9] Pozzan, A. Molecular descriptors and methods for ligand based virtual high throughput screening in drug discovery. *Curr. Pharm. Des.*, **2006**, *12*, 2099-2110.
- [10] Stahura, F.L.; Bajorath, J. New methodologies for ligand-based virtual screening. *Curr. Pharm. Des.*, **2005**, *11*, 1189-1202.
- [11] Dudek, A.Z.; Arodz, T.; Gálvez, J. Computational methods in developing quantitative structure-activity relationships (QSAR): a review. *Comb. Chem. High Throughput Screen.*, **2006**, *9*, 213-228.
- [12] Pérez Nueno, V.; Ritchie, D.W.; Rabal, O.; Pascual, J.; Borrell, J.; Teixido, J. Comparison of ligand-based and receptor-based virtual screening of HIV entry inhibitors for the CXCR4 and CCR5 receptors using 3D ligand shape matching and ligand-receptor docking. *J. Chem. Inf. Model.*, **2008**, *48*, 508-533.
- [13] Hawkins, P.C.D.; Skillman, A.G.; Nicholls, A. Comparison of shape-matching and docking as virtual screening tools. *J. Med. Chem.*, **2007**, *50*, 74-82.
- [14] Chen, H.; Lyne, P.D.; Giordanetto, F.; Lovell, T.; Li, J. On evaluating molecular-docking methods for pose prediction and enrichment factors. *J. Chem. Inf. Model.*, **2006**, *46*, 401-415.
- [15] Bissantz, C.; Schalon, C.; Guba, W.; Stahl, M. Focused library design in GPCR projects on the example of 5-HT(2c) agonists: comparison of structure-based virtual screening with ligand-based search methods. *Proteins*, **2005**, *61*, 938-952.
- [16] Sciabola, S.; Carosati, E.; Baroni, M.; Mannhold, R. Comparison of ligand-based and structure-based 3D-QSAR approaches: a case study on (aryl)-bridged 2-aminobenzonitriles inhibiting HIV-1 reverse transcriptase. *J. Med. Chem.*, **2005**, *48*, 3756-3767.
- [17] Sheridan, R.P.; Kearsley, S.K. Why do we need so many chemical similarity search methods? *Drug Discov. Today*, **2002**, *7*, 903-911.
- [18] Makara, G. Measuring molecular similarity and diversity: total pharmacophore diversity. *J. Med. Chem.*, **2001**, *44*, 3563-3571.
- [19] Schuffenhauer, A.; Gillet, V.J.; Willett, P. Similarity searching in files of three-dimensional chemical structures: analysis of the BIOSTER database using two-dimensional fingerprints and molecular field descriptors. *J. Chem. Inf. Comput. Sci.*, **2000**, *40*, 295-307.
- [20] Matter, H.; Pötter, T. Comparing 3D pharmacophore triplets and 2D fingerprints for selecting diverse compound subsets. *J. Chem. Inf. Comput. Sci.*, **1999**, *39*, 1211-1225.
- [21] Brown, R.D.; Martin, Y.C. The information content of 2D and 3D structural descriptors relevant to ligand-receptor binding. *J. Chem. Inf. Comput. Sci.*, **1997**, *37*, 1-9.
- [22] Evers, A.; Hessler G.; Matter H.; Klabunde T. Virtual screening of biogenic amine-binding G-protein coupled receptors: comparative evaluation of protein- and ligand-based virtual screening protocols. *J. Med. Chem.*, **2005**, *48*, 5448-5465.
- [23] Jones, G.; Willet, P.; Glen, R.C.; Leach, A.R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.*, **1997**, *267*, 727-748.
- [24] Hindle S.; Rarey M.; Buning C.; Lengauer T. Flexible docking under pharmacophore type constraints. *J. Comput. Aided Mol. Des.*, **2002**, *16*, 129-149.
- [25] Catalyst is available from Accelrys, 9685 Scranton Road, San Diego, CA 92121, US. <http://accelrys.com>
- [26] Rarey, M.; Dixon, J. S. Feature trees: a new molecular similarity measure based on tree matching. *J. Comput. Aided Mol. Des.*, **1998**, *12*, 471-490.
- [27] Lemmen, C.; Lengauer, T.; Klebe, G. FLEXS: a method for fast flexible ligand superposition. *J. Med. Chem.*, **1998**, *41*, 4502-4520.
- [28] Schneider, G.; Clement-Chomienne, O.; Hilfiger, L.; Schneider, P.; Kirsch, S.; Bohm, H. J.; Neidhart, W. Virtual screening for bioactive molecules by evolutionary *de novo* design. *Angew. Chem. Int. Ed.*, **2000**, *39*, 4130-4133.
- [29] Schneider, G.; Neidhart, W.; Schmid, G. Scaffold-Hopping by topological pharmacophore search: a contribution to virtual screening. *Angew. Chem. Int. Ed.*, **1999**, *38*, 2894-2896.
- [30] Grethe, G.; Moock, T.E. A new tool for the synthetic chemist. *J. Chem. Inf. Comput. Sci.*, **1990**, *30*, 511-520.
- [31] Jorgensen, W.L.; Duffy, E.M. Prediction of drug solubility from structure. *Adv. Drug Deliv. Rev.*, **2002**, *54*, 355-366.
- [32] Singh N.; Chev e G.; Ferguson D.M.; McCurdy C. A combined ligand-based and target-based drug design approach for G-protein coupled receptors: application to salvinorin A, a selective kappa opioid receptor agonist. *J. Comput. Aided Mol. Des.*, **2006**, *20*, 471-493.
- [33] Zhang, Q.; Muegge, I. Scaffold hopping through virtual screening using 2D and 3D similarity descriptors: ranking, voting, and consensus scoring. *J. Med. Chem.*, **2006**, *49*, 1536-1548.
- [34] Friesner, R.A.; Banks, J.L.; Murphy, R.B.; Halgren, T.A.; Klicic, J.J.; Mainz, D.T.; Repasky, M.P.; Knoll, E.H.; Shelley, M.; Perry, J.K.; Shaw, D.E.; Francis, P.; Shenkin P.S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.*, **2004**, *47*, 1739-1749.
- [35] McGaughey, G.B.; Sheridan, R.P.; Bayly, C.; Culbertson, C.; Kreatsoulas, C.; Lindsley, S.; Maiorov, V.; Truchon, J.F.; Cornell, W.D. Comparison of topological, shape, and docking methods in virtual screening. *J. Chem. Inf. Comput. Sci.*, **2007**, *47*, 1504-1519.
- [36] Baber, J.C.; Shirley, W.A.; Gao, Y.; Feher, M. The use of consensus scoring in ligand-based virtual screening. *J. Chem. Inf. Model.*, **2006**, *46*, 277-288.
- [37] Hu, X.; Prehna, G.; Stebbins, C.E. Targeting plague virulence factors: a combined machine learning method and multiple conformational virtual screening for the discovery of Yersinia protein kinase A inhibitors. *J. Med. Chem.*, **2007**, *50*, 3980-3983.
- [38] ADMET Predictor, SimulationsPlus, Lancaster, CA. <http://www.simulations-plus.com/>.
- [39] Clausen, H.; Buning, C.; Rarey, N.; Lengauer, T. FlexE: efficient molecular docking considering protein structure variations. *J. Mol. Biol.*, **2001**, *308*, 377-395.
- [40] Wang, R.; Lai, L.; Wang, S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput. Aided Mol. Des.*, **2002**, *16*, 11-26.

- [41] Stendl, T.M.; Crump, C.E.; Hayden, F.G.; Langer, T. Pharmacophore modeling, docking, and principal component analysis based clustering: combined computer-assisted approaches to identify new inhibitors of the human rhinovirus coat protein. *J. Med. Chem.*, **2005**, *48*, 6250-6260.
- [42] Cerius, Accelrys Inc: San Diego, CA; <http://www.accelrys.com>
- [43] Ilib diverse, Inte: Ligand GmbH: Maria Enzersdorf, Austria, <http://www.inteligand.com>
- [44] Rastelli, G.; Pacchioni, S.; Sirawaraporn, W.; Sirawaraporn, R.; Parenti, M.D.; Ferrari, A.M. Docking and database screening reveal new classes of Plasmodium falciparum dihydrofolate reductase inhibitors. *J. Med. Chem.*, **2003**, *46*, 2834-2845.
- [45] Rastelli, G.; Sirawaraporn, W.; Sirawaraporn, R.; Vilaivan, T.; Kamchonwongpaisan, S.; Quarrell, R.; Lowe, G.; Thebtaranonth, Y.; Yuthavong, Y. Interaction of pyrimethamine, cycloguanil, WR99210 and their analogues with Plasmodium falciparum dihydrofolate reductase: structural basis of antifolate resistance. *Bioorg. Med. Chem.*, **2000**, *8*, 1117-1128.
- [46] Gschwend, D.A.; Kuntz, I.D. Orientational sampling and rigid-body minimization in molecular docking revisited: on-the-fly optimization and degeneracy removal. *J. Comput. Aided Mol. Des.*, **1996**, *10*, 123-132.
- [47] Zhang, Q.Y.; Wan, J.; Xu, X.; Yang, G.F.; Ren, Y.L.; Liu, J.J.; Wang, H.; Guo, Y. Structure-based rational quest for potential novel inhibitors of human HMG-CoA reductase by combining CoMFA 3D QSAR modeling and virtual screening. *J. Comb. Chem.*, **2007**, *9*, 131-138.
- [48] Abrahamian, E.; Fox, P.C.; Naerum, L.; Christensen, I.T.; Thøgersen, H.; Clark, R. D. Efficient generation, storage, and manipulation of fully flexible pharmacophore multiplets and their use in 3-D similarity searching. *J. Chem. Inf. Comput. Sci.*, **2003**, *43*, 458-468.
- [49] Moro, S.; Braiuca P.; Deflorian F.; Ferrari C.; Pastorin G.; Cacciari B.; Baraldi P. G.; Varani K.; Borea P.A.; Spalluto G. Combined target-based and ligand-based drug design approach as a tool to define a novel 3D-pharmacophore model of human A3 adenosine receptor antagonists: pyrazolo[4,3-e]1,2,4-triazolo[1,5-e]pyrimidine derivatives as a key study. *J. Med. Chem.*, **2005**, *48*, 152-162.
- [50] Molecular Operating Environment (MOE 2003.02), C.C.G., Inc., 1255 University St., Suite 1600, Montreal, Quebec, Canada, H3B3X3, **2003**.
- [51] Luzhkov, V.B.; Selisko, B.; Nordqvist, A.; Peyrane, F.; Decroly, E.; Alvarez, K.; Karlen, A.; Canard, B.; Aqvist, J. Virtual screening and bioassay study of novel inhibitors for dengue virus mRNA cap (nucleoside-2'O)-methyltransferase. *Bioorg. Med. Chem.*, **2007**, *15*, 7795-7802.
- [52] Cherkasov, A.; Ban, F.; Li, Y.; Fallahi, M.; Hammond, G.L. Progressive docking: a hybrid QSAR/docking approach for accelerating *in silico* high throughput screening. *J. Med. Chem.*, **2006**, *49*, 7466.
- [53] Cherkasov, A.; Shi, Z.; Fallahi, M.; Hammond, G. Successful *in silico* discovery of novel nonsteroidal ligands for human sex hormone binding globulin. *J. Med. Chem.*, **2005**, *48*, 3203-3213
- [54] Klön, A.E.; Glick, M.; Thoma, M.; Acklin, P.; Davies, J.W. Finding more needles in the haystack: A simple and efficient method for improving high-throughput docking results. *J. Med. Chem.*, **2004**, *47*, 2743-2749.
- [55] Jacobsson, M.; Liden, P.; Stjærnschantz, E.; Bostrom, H.; Norinder, U. Improving structure-based virtual screening by multivariate analysis of scoring data. *J. Med. Chem.*, **2003**, *46*, 5781-5789.
- [56] Klön, A.E.; Glick, M.; Davies, J.W. Application of machine learning to improve the results of high-throughput docking against the HIV-1 protease. *J. Chem. Inf. Comput. Sci.*, **2004**, *44*, 2216-2224.
- [57] Klön, A.E.; Glick, M.; Davies, J.W. Combination of a naive Bayes classifier with consensus scoring improves enrichment of high-throughput docking results. *J. Med. Chem.*, **2004**, *47*, 4356-4359.
- [58] Yoon, S.; Smellie, A.; Hartsough D.; Filikov A. Surrogate docking: structure-based virtual screening at high throughput speed. *J. Comput. Aided Mol. Des.*, **2005**, *19*, 483-497.
- [59] Parrill, A. L.; Echols, U.; Nguyen, T.; Pham, T.-C.; Hoeglund, A.; Baker, D.L. Virtual screening approaches for the identification of non-lipid autotaxin inhibitors. *Bioorg. Med. Chem.*, **2008**, *16*, 1784-1795.
- [60] Dragon, Milano Chemometrics, Milano, Italy. <http://www.vclab.org/lab/edragon/>.
- [61] Gavernet, L.; Talevi, A.; Castro, E.A.; Bruno-Blanch, L.E. A combined virtual screening 2D and 3D QSAR methodology for the selection of new anticonvulsant candidates from a natural product library. *QSAR Comb. Sci.*, **2008**, *27*, 1120-1129.
- [62] Guba, W.; Neidhart, W.; Nettejoven, M. Novel and potent NPY5 receptor antagonists derived from virtual screening and iterative parallel chemistry design. *Bioorg. Med. Chem. Lett.*, **2005**, *15*, 1599-1603.
- [63] Mukherjee, P.; Desai, P.; Ross, L.; White, E.L.; Avery, M.A. Structure-based virtual screening against SARS-3CLpro to identify novel non-peptidic hits. *Bioorg. Med. Chem.*, **2008**, *16*, 4138-4149.
- [64] Desai, P.V.; Patny, A.; Gut, J.; Rosenthal, P.J.; Tekwani, B.; Srivastava, A.; Avery, M. Identification of novel parasitic cysteine protease inhibitors by use of virtual screening. 2. The available chemical directory. *J. Med. Chem.*, **2006**, *49*, 1576-1584.
- [65] Desai, P.V.; Patny, A.; Sabnis, Y.; Tekwani, B.; Gut, J.; Rosenthal, P.; Srivastava, A.; Avery, M. Identification of novel parasitic cysteine protease inhibitors using virtual screening. 1. The Chem-Bridge database. *J. Med. Chem.*, **2004**, *47*, 6609-6615.
- [66] Polgar, T.; Keseru, G.M. Virtual screening for beta-secretase (BACE1) inhibitors reveals the importance of protonation states at Asp32 and Asp228. *J. Med. Chem.*, **2005**, *48*, 3749-3755.
- [67] Evers, A.; Klabunde, T. Structure-based drug discovery using GPCR homology modeling: successful virtual screening for antagonists of the alpha1A adrenergic receptor. *J. Med. Chem.*, **2005**, *48*, 1088-1097.
- [68] Maiorov, V.; Sheridan, R.P. Enhanced virtual screening by combined use of two docking methods: getting the most on a limited budget. *J. Chem. Inf. Model.*, **2005**, *45*, 1017-1023.
- [69] Miller, M.D.; Kearsley, S.K.; Underwood, D.J.; Sheridan, R.P. FLOG: a system to select 'quasi-flexible' ligands complementary to a receptor of known three-dimensional structure. *J. Comput. Aided Mol. Des.*, **1994**, *8*, 153-174.
- [70] ICM-Pro, Molsoft, LLC: LaJolla, CA. http://www.molsoft.com/icm_pro.html
- [71] Miteva, M.A.; Lee, W.H.; Montes, M.O.; Villoutriex, B.O. Fast structure-based virtual ligand screening combining FRED, DOCK, and Surflex. *J. Med. Chem.*, **2005**, *48*, 6012-6022.
- [72] McGann, M.; Almond, H.; Nicholls, A.; Grant, J.A.; Brown, F. Gaussian docking functions. *Biopolymers*, **2003**, *68*, 76-90.
- [73] Jain, A.N. Hierarchical database screenings for HIV-1 reverse transcriptase using a pharmacophore model, rigid docking, solvation docking, and MM-PB/SA. *J. Med. Chem.*, **2003**, *46*, 499-511.
- [74] Wang, J.; Kang, X.; Kuntz, I.D.; Kollman, P.A. *J. Med. Chem.*, **2005**, *48*, 2432-2444.
- [75] Nicholls, A. What do we know and when do we know it? *J. Comput. Aided Mol. Des.*, **2008**, *22*, 239-275.